UNIVERSITY OF PENNSYLVANIA

A STUDY OF THE COMPARATIVE VALUE OF TWO METHODS OF IMPROVING PROBLEM SOLVING ABILITY IN ARITHMETIC

ROY EDGAR ADAMS

A THESIS

IN EDUCATION

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

PHILADELPHIA
1930

A STUDY OF THE COMPARATIVE VALUE OF TWO METHODS OF IMPROVING PROBLEM SOLVING ABILITY IN ARITHMETIC

An Experiment with Public School Pupils of Grades Three and Four to Determine the Relative Value of Two Methods of Instruction in Solving Problems in Arithmetic, with Special Consideration of the Effect of Detailed Analysis Upon Pupils' Problem Solving Ability

---

A THESIS

IN EDUCATION

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF PENNSYLVANIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

ROY EDGAR ADAMS

PHILADELPHIA
1930

R. UNIV.
BIBLIOTHEEK
LEIDEN

ACKNOWLEDGMENTS

The writer gratefully acknowledges the guidance of Dr. Francis M. Garver of the University of Pennsylvania, under whose encouraging supervision this experiment was made possible. He is also grateful to Dr. LeRoy A. King and Dr. F. Herrick Conners, who also were members of the faculty committee of the University, for their helpful criticism and encouragement. To the many principals and teachers of Philadelphia and Reading, Pa., acknowledgment is made for their willing supervision and co-operative participation which led to the completion of this experimental study. Mr. Landis Tanger, then Superintendent of Schools of Reading, Pa., was exceedingly kind in his offering of the facilities of the Reading schools for the experimental work in that city, and the grade supervisors, Miss Rapp and Miss Dowling, were very helpful in the conduct of the work there.

Special mention must be made of the co-operation of the officials, principals and teachers in the Philadelphia public school system. Associate Superintendent Dr. Armand J. Gerson very generously permitted the use of the schools of Philadelphia, and the Division of Research was always ready with pertinent advice and criticism in matters of administration and technical procedure.

Sincere thanks are due the members of the Seminar in Elementary Education at the University of Pennsylvania during the years 1928–30 for their valuable personal assistance as well as for many helpful suggestions.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547325_0004

TABLE OF CONTENTS

I

INTRODUCTION 7
a. The Status of Problem Solving 7
b. Statement of the Problem 10

II

ORGANIZATION OF THE PROBLEM TESTS 12
a. The Preliminary Tests 13
b. Preliminary Testing and Evaluation 16
c. The Final Tests 16
d. Correlation of Test 1 and Test 2 18

III

DEVELOPMENT OF THE EXPERIMENTAL PROCEDURE 20
a. The Philadelphia Experiment of 1927–28 20
1. The Two Methods of Instruction 20
2. Schools and Types of Classes 21
3. Experiment Results 22
b. The 1929 Experiment in Reading, Pa. 23
1. Changes in Previous Method 23
2. Selecting Experimental Classes 24
3. The Control Method in Reading, Pa. 27
4. Teacher Training and Experience 29
5. Supervision 29
6. Experiment Results 31

IV

THE FINAL PHILADELPHIA EXPERIMENT, 1929 34
a. Schools and Classes Selected—Comparisons 34
1. Intelligence and Problem Solving Ability 40
b. Experiment Data 43
c. The Two Methods of Instruction 45
d. Teacher Training and Experience 46
e. Supervision 48
f. Results of the Experiment—Tables and Graphs 50
g. Comparison of Results and Conclusion 53

APPENDIX 59
a. The Test Scales 59
b. Test Directions 60
c. The Analysis Method 61
d. The Control Method 63
e. Median Scores of Classes 65
f. Evaluation of Tests A and B 67

BIBLIOGRAPHY 68

I

Introduction

Arithmetic instruction seems to have two well defined aims: the formation of habits and skills in the manipulation of numbers, and the development of a kind of skill in reasoning. To realize the first, the child is taught and carefully drilled in the four fundamental operations, with integers, fractions, and decimals, instructed in the method of handling per cents, and given some knowledge of involution and evolution. The hoped-for goal is complete, automatic mastery of number combinations, with the acquisition of skill and accuracy in handling numbers in all processes. The "why" is not so much stressed as the "how", and perfection in this phase of arithmetic is somewhat similarly regarded as perfection in the use of tools. Without proper ability to handle the tools, no task can be completed satisfactorily. This side of arithmetic has often been called the "habit" side.

But many of the actual arithmetic situations that are met in everyday life require more than the mere knowledge of "how to do"—they require that the individual shall decide "what to do." These situations present themselves in the form of problems in which certain data are given and an unknown fact is to be determined. It is here that the individual must "make up his mind" as to *what* he shall do to get the required answer. At this point real reasoning enters. The solving of arithmetic problems, therefore, constitutes the reasoning side of arithmetic.

The term "problem" is used here in contrast with the term "example". In an example in arithmetic the process to be used is indicated, viz: $5+6=?$ $33-27=?$ $25\times8=?$ etc. It may be expressed in words, as: Add twenty-five to the sum of nine and fourteen. However, as long as the process or processes to be used are indicated, by arithmetical sign or by word, the situation is to be known as an example. The problem, on the other hand, expressed either orally or in written form, does not indicate by sign or word the process required for solution. The solver must *decide* for himself the process to be used, and in making this decision he is solving a problem. To illustrate, consider the

following: "How much money must I have to buy 3 cans of corn at 18 cents a can?" There is nothing in the foregoing to show that the answer will be obtained by multiplication. The numbers 3 and 18 might be added, subtracted, multiplied or divided. It becomes a problem by the very presence of the necessity for choice of operation.

Teachers of arithmetic are giving more and more thought to this side of their work. Neglected and often ignored, the reasoning side of arithmetic (the solving of problems) has not heretofore met with much consideration. Inquiries have recently been made and experiments conducted with a view to determining the best method of improving problem solving ability.

Dr. F. S. Knight, in the preface to the University of Iowa Monograph on Education, first series, number 6, entitled "An Evaluation of Three Techniques for Improving Ability to Solve Arithmetic Problems", by Dr. O. S. Lutes, points out the fact of our relative lack of knowledge about problem solving instruction. He believes that any serious study, supported by solid experimental evidence will be well received by those working in this field. He suggested, among others, these questions:

1. What techniques should we, as teachers, use in our problem work?
2. Should we use set forms of solutions in teaching pupils?
3. Shall we teach the pupil to develop a "system" of problem attack? Or shall we give him problems and let him "sink or swim"?

There have been a few attempts to answer these questions, as noted later, but the final answer, if such there is, has probably not as yet been found. As Dr. Lutes states, "it is probably true that less is actually known about how to teach this (problem solving) skill than any other important aspect of elementary school arithmetic." (*Op. cit.*, p. 7.) Buswell and Judd[1] show us that very few pages of educational material are devoted either to problems or to the study of arithmetical reasoning. It is probably true that there is little on record that shows a deliberate attempt to improve problem solving ability.

[1] Buswell, Guy T. and Judd, Chas. H., "Summary of Educational Investigation Relating to Arithmetic", University of Chicago, 1925.

One of the quite generally recognized causes of lack of success in problem solving has been poor reading comprehension. To determine the effect of improvement in reading comprehension on the ability to solve arithmetical problems, several educators have made valuable contributions as the result of scientific investigation. Wilson[2] gave reading drills on the meaning of problems and in arithmetical language. She found that this produced considerable gain in scores on the Stone's Reasoning Tests. In somewhat the same manner, Greene[3] showed that directed reading of arithmetic problems would secure improvement in problem solving scores. Lessenger[4] found that drill in silent reading improved the results in problem solving tests. Osburn[5] analyzed the errors and possible remedial measures. A very recent study was made by V. E. Chase[6] of the Bureau of Research of Fordson, Michigan, a case study of seventeen pupils in an attempt to find causes of error and possible remedial measures.

Several others have made studies of pupils' inability to reason and of the factors involved in problem solving. Arithmetic textbooks have been carefully examined to determine the extent to which they are at fault and contribute to the general poor results. Much has been done in the field of standard tests, and we consequently now have Sone's Reasoning Tests, Monroe's Reasoning Test, Buckingham's Problem Analysis Test, Stevenson's Problem Analysis Test and the Stanford Achievement Tests, Part 5. Many more recent tests have been formulated.

One of the recent investigations into method was made by

---

[2] Wilson, Estaline, "Improving the Ability to Read Arithmetic Problems". *Elementary School Journal*, XXII, pp. 380-86.

[3] Greene, Harry A., "A Directed Drill in the Comprehension of Verbal Problems in Arithmetic", *Journal of Educational Research*, X (Nov. 1925), pp. 297-300.

[4] Lessenger, Waldo E., "Motivation and the Accomplishment Quotient Technique," College of Educ. Library, University of Iowa, 1925.

[5] Osbourn, W. J., "Diagnosis and Remedial Treatment for Errors in Arithmetical Reasoning", State Department of Pur. Instr., Madison, Wis., 1922-mimeo. mater.

[6] Chase, V. E., "Diagnosis and Treatment of Some Common Difficulties in Solving Verbal Problems in Arithmetic", Bu. 12, Bureau of Res., Fordson, Mich.

Washburne and Osborne[7] who in an experimental way endeavored to find the relative value of three different methods of problem solving. The conclusion reached, that no method of analysis or special technique was capable of producing noticeably greater improvement, has generally been accepted up to the present time. Along a similar line, Dr. Lutes[8] experimented with three different techniques, or methods, and his results are recorded in the University of Iowa Monograph mentioned above. Briefly summarized his conclusions were:

1. "All pupils of normal intelligence can profit from definite instruction in problem solving.
2. The method that emphasizes computation skill produces the greatest gain in both reasoning and computation.
3. Motivation is an important factor in securing improvement in problem solving ability. . . .
4. Improvement in computation accuracy does increase ability to solve problems; whether it increases ability to do arithmetical reasoning cannot be stated from the results of the study. . . ."

In both of these investigations the children used were in the higher elementary grades. Dr. Lutes used 6B classes, and Washburne and Osborne used 760 pupils of the sixth and seventh grades.

It is now customary in most school systems to begin the definite work of problem solving in the lower elementary grades, often as low as the second year. Little or nothing has been done in an experimental way with these children. They are at the point of first impressions, at the stage when success in problem solving may be won or lost. What shall be done with children in these early grades to give them a successful start in this difficult field of arithmetic?

The Problem

An investigation was made by the writer to determine the effect of a specific method of reasoning analysis in solving

---

[7] Washburne, C. W., Osborne, R., "Solving Arithmetic Problems." *El. Sch. Journal*, Dec. 1926, pp. 296-304.

[8] *Op. cit.*, p. 38.

problems in the third and fourth grades, in the public schools of Philadelphia and Reading, Pa. The entire study covered a period of three years and was begun in Philadelphia schools in October, 1928. The first year's work included the formation of two equivalent problem tests and a ten-weeks' period of experimentation. The results are indicated in later pages. In a slightly modified form the experiment was repeated, this time in Reading, Pa., during the spring of 1929, and the results of the work in that city are also stated later. Again, in 1929–30, the experiment was made in Philadelphia, under carefully controlled conditions, and the results of that work constitute the main part of this report.

The purpose of this study is to record the evidence, experimentally secured, bearing on the relative merits of two different methods of improving problem solving ability in the third and fourth grades. The child confronted with a problem feels that he must "do something". Shall he guess? Shall he follow out a set line of reasoning? Shall he be trained in the use of a detailed method of analysis and in the use of special definite helps and devices?

In Philadelphia, the Course of Study in Arithmetic specifically indicates a method of problem teaching. It is the beginning, at least, of a formal analysis method. Would it be improved by an extension to a more detailed analytic form and by the addition of certain special helps and devices?

To determine the answer to these questions, an experiment was conducted in an actual public school situation, on a wide scale, with definite methods of procedure set down. In the same manner as Dr. Lutes' study, mentioned above, it "is an attempt to discover desirable techniques for improving mass instruction, that is, to discover the worth of techniques usable in classroom situations." Although it differs from that study in that it is concerned with younger children, the results of the experiment made in connection with this study agree in many respects with the findings of Lutes, and with the conclusions of Washburne and Osborne.

II

Organization of the Problem Tests

A. In order to have objective records of progress attained it was necessary to have two problem tests, of as nearly equal difficulty as possible, for a preliminary and final testing. Rather than use any of the prevailing standard tests generally recognized, which did not seem adequately to meet the requirements of the problem solving work in Philadelphia, for grades 3 and 4 particularly, it was decided that two tests should be devised. These tests, to be known as Test A and Test B, were prepared to include samples of all possible problem situations within the course of study limits of grades 3 and 4 in the Philadelphia public schools. They were paired for similarly in form of difficulty and tentatively arranged in progressive order of difficulty. They were then subjected to preliminary testing in several schools that were not to be included in the experiment in order to establish equality of the problems in the two tests and proper order for progressive difficulty arrangement.

These tests, each containing sixteen representative problems, in the order that was established by the preliminary testing in Philadelphia, offered the following problem situations:

1. Addition: finding the total cost of two articles.
2. Subtraction: finding a remainder when a part is taken away.
3. Subtraction: do.—larger numbers.
4. Addition: finding the total cost of two articles.
5. Multiplication: finding a number that is x times another number.
6. Multiplication: finding cost of x things at c cents each.
7. Subtraction: finding the amount needed to equal a given number, with the subtrahend stated first.
8. Division: divisor and dividend the same kind of things.
9. Division: sharing equally.
10. Subtraction: finding the amount needed to equal a given number, with the minuend stated first.
11. Division: finding a fractional part of a number.
12. Multiplication: involving U. S. money, with two-place multiplier and multiplicand.
13. Division: dividend and quotient of the same kind of things.
14. Two-step problem: finding an "average".

15. Two-step problem: finding total cost and making change.
16. Two-step problem: installment buying.

The problems in these tests were marked correct if the right fundamental operation was used for their solution. No attention was paid to accuracy of mechanics of procedure. Throughout this discussion correct solution is to be understood as meaning correct choice of operation. The teachers who administered these tests and marked the pupils' papers were instructed in this part of the procedure in the Test Directions (see page 60 of the Appendix) which were given to them.

The tests were administered first to 618 pupils of grades 3A-4B inclusive in several Philadelphia public schools.

As the two tests were given in the schools with but a two-day interval, it was thought that the element of recency might have influenced the results which show generally better scores for Test B. Accordingly, in another school, the tests were again given, but in reverse order, Test B being given first and followed in two days by Test A. Again the second test, this time Test A, showed slightly better results throughout. However, these preliminary testings showed the necessity for certain adjustments, and when these had been made, both tests were again tried out with a new set of pupils in other schools. On the basis of 1033 pupils tested, the number of pupils correctly reasoning each problem is shown in the table on the page 15.

PRELIMINARY TEST—A

1. How much money must Alice take to the store to buy a 7-cent box of crackers and a can of peas that cost 18 cents?

2. A fruit man had 36 oranges in a box. He sold 24 of them to Tom's mother. How many oranges were left in the box?

3. A teacher had 58 reading books in her closet. She sent 15 of them to the teacher in the next room. How many did she have left?

4. A grocer sold a can of peas to Mrs. Thomas for 18 cents and a basket of potatoes for 38 cents. How much money should he receive from Mrs. Thomas?

5. The postman told Frank that he walked 6 miles each day in delivering his mail. How far would the postman walk in 8 days?

6. How much money will Harry need to buy 4 wheels for his roller skates if the wheels cost 22 cents each?

7. Mary saw a doll's dress in a store window. It cost 80 cents. How much more money does she need to buy it if she has only 62 cents in her bank?

8. How many piles of oranges could a storekeeper make from a box of 54 oranges if he put 6 oranges in a pile?

9. Four boys earned 48 cents selling old newspapers. If they share the money equally, how much will each boy get?

10. Helen has saved $3.25. How much more money does she need to buy a doll that costs $6.00?

11. John puts ¼ of all the money he earns into his savings bank. Last week he earned 36 cents. How much should he put into his bank?

12. Bob earns $.25 an hour working at a grocery store. If he should work 13 hours this week, how much money will he receive?

13. A clothing dealer paid $60.30 for 9 suits of clothes to sell in his store. What was the average cost of each suit?

14. On the first three days of an automobile trip, Tom and his father traveled 92 miles, 87 miles, and 97 miles. What was the average distance traveled each day?

15. Dorothy's mother bought a 6-pound roast of beef at $.32 a pound. She gave the storekeeper $5.00. How much change should she receive?

16. Mr. Jamison bought an automobile that cost $1050. He made a first payment of $450 and agreed to pay $75 a month until the rest of the money was paid. How long did it take him to pay the entire cost?

FINAL TEST—B.

1. Harry bought a top for 8 cents and a ball for 18 cents. How much did both things cost him?

2. Howard had 28 cents in his bank. He bought a top that cost 10 cents. How much did he have left?

3. Harry bought a bag of marbles. The bag contained 36 marbles. If he gave 15 of them to his little brother, how many would Harry have left?

4. Jane's mother paid 35 cents for a bag of flour and 28 cents for oranges. How much money did she give the grocer?

5. Dorothy put 4 pennies into her bank each day. How many pennies will she put in in 9 days?

6. How much money will Jane have to pay for 3 pounds of nuts if the nuts cost 28 cents a pound?

7. Frank wants to buy a baseball bat that costs 75 cents, but he has only 53 cents in his bank. How much money does he need?

8. How many days will it take a class to read a story that is 36 pages long, if the class reads 4 pages a day?

9. Three boys bought a baseball that cost 75 cents. How much should each boy pay if they agree to share the cost equally?

10. Joe has saved $2.50 to buy a catcher's glove, but the glove costs $3.75. How much more money will Joe have to have?

11. Harry's father gives him ⅓ of all the money he receives from selling the old newspapers in his house. In one week they received 24 cents for the newspapers they sold. How much of the money received was Harry's?

12. How much will Fred's mother have to pay for a turkey that weighs 14 pounds and costs $.53 a pound?

13. A farmer received $13.20 for 8 bushels of potatoes. What was the average price received for them by the bushel?

14. Harry's marks in his tests last week were 85, 92, and 90. What was his average mark for the week?

15. How much change should you receive from $2.00 if you buy a 5-pound roasting chicken at $.37 a pound?

16. A farmer bought a horse for $165 paying $40 down and the balance in 5 months. How much did he pay each month?

TABLE 1

COMPARISON OF TESTS A AND B, 1033 PUPILS

PROBLEM	TEST A	TEST B
1	873	855
2	840	878
3	805	792
4	657	662
5	744	808
6	642	677
7	660	616
8	463	440
9	475	398
10	386	385
11	318	425
12	293	266
13	158	165
14	87	177
15	136	111
16	57	47
	7594	7672

The foregoing results show the advisability of certain changes in order to have the test arranged in order of progressive difficulty, on the basis of the number of pupils tested. However they were considered satisfactory for a complete preliminary try-out in the first Philadelphia experiment.

The tests at this stage of their development are here shown:

At the end of the first Philadelphia experiment the two tests were again compared, using the results obtained from the 2321 pupils[1] included in that experiment. The larger number of pupils shows about the same relative results and indicates the necessity for additional adjustments before the tests could be considered satisfactory in arrangement for progressive difficulty. The procedures involved to accomplish the necessary adjustments were:

1. Change of certain problems in their position in the tests.

2. Elimination of certain problems. The results on some problems were so close as to indicate that it was unnecessary to include all of them, even though they did require different processes for solution.

3. Addition of a few more difficult problems to the tests. These problems were designed to be of such difficulty that only an exceptional fourth grade pupil could solve them. Three separate testings, involving 1300 pupils, were used to select finally these extra problems and to determine their position in the tests.

4. Reduction of the tests to 12 problems each. Certain problems were eliminated as being unnecessary for the purpose. This reduction and a rearrangement in order of difficulty to conform more closely to the results obtained from the try-out testings produced the tests in the form shown on pages 59 and 60 of the appendix. In that form they were used in the experimental work in Reading, Pa., and in the final Philadelphia experiment.

Final Tests

As now constituted, the tests include the following problem situations in each form:

1. *Subtraction:* finding remainder when a part is taken away.
2. *Addition:* finding the total cost of two articles.
3. *Subtraction:* finding amount needed to equal another number, with the subtrahend stated first.
4. *Multiplication:* finding cost of *X* articles at *O* cents each.
5. *Division:* sharing equally.

---

[1] See page 61 of appendix.

6. *Division:* divisor and dividend the same kind of thing.
7. *Multiplication:* involving money.
8. *Two-step problem:* finding cost by multiplication and making change.
9. *Two-step problem:* finding cost of dissimilar things and making change.
10. *Two-step problem:* finding an "average".
11. *Two-step problem:* installment buying.
12. *Three-step problem:* special difficulty in finding a fractional part.

The results on these tests in Reading, Pa., showed that the same progressive order of difficulty did not hold for that city. Possible curricular differences may account for this. It was particularly noticeable that problem 7, a multiplication situation involving money, which ranked seventh in Philadelphia, is only fifth in difficulty in Reading. The differences between the two results, however, are consistent in the two tests, and the following table shows the number of pupils of Reading who correctly reasoned each problem in each of the test forms:

TABLE 2

COMPARISON OF TEST A AND TEST B IN READING, PA., IN NUMBER OF PUPILS CORRECTLY REASONING EACH PROBLEM

PROBLEM	TEST A (1746 PUPILS)	TEST B (2102 PUPILS)
1	1463	1925
2	1184	1647
3	1279	1804
4	1150	1733
5	759	1394
6	574	1280
7	1074	1473
8	296	675
9	243	485
10	219	645
11	58	228
12	24	125

In both tests problem 7 was much easier for the Reading pupils to reason than for the pupils in Philadelphia. It is to be noted, also, that the addition situation of problem 2 in the second test is more difficult relatively for Reading pupils than the subtraction situation in problem 3 or the multiplication situation

in problem 4. The large increase in the number of pupils who correctly reasoned problem 10, finding an "average", was probably due to definite teaching of this kind of problem situation on the part of teachers who may have expected a similar problem to appear in the second test. On the whole, however, the results in Reading, Pa., show the tests to be of about the same progressive difficulty as was determined by the Philadelphia experimentation.

To determine the validity of the tests as comparable measuring instruments they were administered to 572 pupils of grades 3A to 4B inclusive in four Philadelphia schools. By the use of the product-moment formula as developed by the Otis Correlation Chart, the coefficient of correlations was found to be .842 ± .007. The scores were then sorted according to the grade placement of the 572 pupils, and separate correlations were determined for each grade. These correlations are shown in the following table:

TABLE 3
CORRELATION OF TEST 1 AND TEST 2, DETERMINED FROM 572 PUPILS OF GRADES 3A–3B IN FOUR PHILADELPHIA SCHOOLS

GRADE	NUMBER OF PUPILS	CORRELATION
Total Group	572	.842 ± .007
4B	184	.848 ± .053
4A	171	.822 ± .034
3B	112	.600 ± .034
3A	105	.530 ± .027

The total group shows that a very high correlation exists between the two tests. As can be expected from smaller numbers of pupils in the separate grades, the coefficients of correlation vary from this total group. It is noted that grades 4A and 4B show about the same high coefficients, but in grades 3B and 3A there is a drop to .600 and .530 respectively. There still remains, however, a very high reliability in the results in these grades. In any consideration of error due to inaccuracy of tests, T. L. Kelley[2] maintains that a reliability coefficient of .50 or more is

[2] Kelley, T. L., "Interpretation of Educational Measurements", p. 211.

satisfactory for *group* measurement. He states, "We shall conclude that a reliability coefficient, when determined from a single grade range of .50 or higher, is demanded of a test which is to be used for group measurement purposes." The reliability coefficient is the correlation between two comparable tests.

From these results we may conclude that the two tests were reliable measuring instruments for use in the grades for which they were devised. They agree with the requirements for the construction of two comparable tests as laid down by Kelley[3]: (1) "the elements of the first test should be as similar in difficulty and type to those in the second, pair by pair, as possible, but (2) should not be so identical in word or form as commonly to lead to memory transfer or correlation between errors."

---

[3] Kelley, T. L., "Statistical Method", p. 203.

III

Development of the Experimental Procedure

The development of the problem tests marked the end of one of the main parts of the experimental program. This development was not attained until two preliminary experimental uses of the tests had been made, first in Philadelphia in 1927-28, and later in Reading, Pa., in the spring of 1929. These preliminary experimentations showed the necessity for several changes in experimental procedure as well as in the tests, and in order that the final methods may be better understood, a brief summary of those experiments is included as a part of the report.

A. The Philadelphia Experiment of 1927-28

Upon the completion of the first organization of the two problem tests, consisting then of 16 problems each, they were used to measure the relative efficiency of two methods of improving problem solving ability in an experiment conducted in grades 3A, 3B, 4A, and 4B in ten Philadelphia public schools.

The two methods of instruction used were known as the Experimental Method and the Control Method; they are shown in detail on pages 61-64 of the Appendix. All the teachers concerned in the work were met in group conference by the one in charge of the experiment, shown the methods to be used, and instructed in the details necessary for uniform control. To each teacher were given copies of General Directions and Test Directions at the time the tests were given. The problems comprising the two tests, Preliminary and Final, copies of which are shown on page 59 of the Appendix, were placed on mimeographed sheets and a copy given to each pupil. The pupils solved as many of the problems as they could on a sheet of paper of the kind regularly used for arithmetic work in the class.

After the Preliminary Test had been administered, score sheets were obtained from the teachers which showed the number of pupils who correctly reasoned each problem. The data were used to check the efficiency of the tests as to their equality and relative position of the problems. The class medians were

not given by the teachers but were determined by the experimenter from a study of the pupils' papers.

The schools selected for the experimental work were chosen as being typical of the various social, economic, and racial sections of the city, selected to obtain as nearly as possible equal pairs in communities of differing types. They were in charge of Principals who were willing to share in the experimental work and who would properly give it the needed supervision. In addition to the two groups of schools used in the experiment, three schools were chosen to act as an "uncontrolled" group as a check on the other groups. As this work in Philadelphia at this time was but preliminary to the final experiment two years later, it is considered sufficient to give but a brief summary of the procedure.

There were three types of classes: the Experimental, that were taught by a detailed method of analysis which is described on page 61 of the Appendix; the Control, which were taught by the method prescribed by the Philadelphia Course of Study in Arithmetic; and the Uncontrolled, the classes that were not directly concerned in the experiment. It is assumed that they were taught by the methods usual to the teachers in charge, and they were not supplied with any information or directions as to method, number of lessons, or program of lessons. As far as they were concerned, no experiment was in progress and they therefore represent the problem solving work that was undertaken under usual conditions and during the experimental period. They were given the first test, at the same time as the other two groups, and later the final test without previous notice that a second test was to be given.

There were 2113 pupils in this preliminary experiment, under the direction of 62 teachers:—experimental, 834 pupils with 24 teachers; control, 772 pupils with 24 teachers; uncontrolled, 507 pupils with 14 teachers.

The work lasted for eight weeks, ending with a final test for which median scores were again obtained. As a result of the definite problem instruction, all classes made a gain in median scores. The gains made by the three types of classes differ quite considerably, however. The experimental classes

that used the analysis method made the greatest gains in grades 3A and 3B. In the 4A and 4B grades, the greatest gains were made by the uncontrolled classes, but in these grades, as in the third, the experimental classes made greater gains than the control classes.

TABLE 4

GAINS IN MEDIAN SCORES—NUMBER OF PROBLEMS CORRECTLY REASONED

GRADE	EXPERIMENTAL	CONTROL	UNCONTROLLED
3A	**3.11**	2.33	1.31
3B	**2.79**	2.45	.71
4A	2.40	.86	**2.76**
4B	2.19	1.79	**2.58**

It should be noted in the above results that the great gains made by the uncontrolled classes in grades 4A and 4B were probably due to the low median scores of these groups in the preliminary test. In a test of progressive difficulty it would undoubtedly be easier to make a large gain at the lower end of the test than at the upper end which contained the more difficult problems. In both of the fourth grades the uncontrolled classes had the lowest medians in the first test, and in spite of their great gains, these classes remain with the lowest medians in the final test. They did not reach the same level of efficiency in problem solving that the other two groups of classes attained.

In the 3A grade the experimental method produced by far the best results. These classes started with a lower median than the control classes, and at the end of the experiment had attained the highest median of the three groups. From these results we may conclude that the experimental was the better method for 3A pupils.

The highest median in the 3B grade was made by the control classes, but the experimental group made more *gain* in median scores during the experimental period. From the standpoint of *improving* results, the experimental method was more efficient in this grade.

In the 4A grade the experimental classes show a marked

superiority over the control classes. They began with a median very much lower than the control classes and practically equalled the median of the control classes in the final test. It is noticed that the uncontrolled classes in this grade made the greatest gain in median scores, but they started with a very low median in the first test. Their gain of 2.76 was the third highest gain made by any of the groups in the experiment.

The experimental group again exceeded the gain of the control group in the 4B grade. It started with a very much lower median in the first test, but although it made a greater improvement, it did not reach the final median attained by the control group. As in the 4A grade, the uncontrolled group again made the greatest gain, but their median in the first test was very low and it was still the lowest of all the groups at the end of the experimental period.

Summary: as the purpose of the experiment was to determine the relative efficiency of two methods (experimental and control) for improving problem solving ability, it must be our conclusion, based on the results obtained from these 2113 pupils, that the experimental method was superior. It produced greater gains in each of the four classes than did the control method. Although the final scores attained by the experimental classes were the highest in only one grade, the 3A, yet they showed greater *improvement* in each of the grades tested.

B. The 1929 Experiment in Reading, Pa.

1. The results of the Philadelphia experiment showed the necessity for a few changes in test material, methods of procedure, and in control of the experimental program.

(a) The tests previously used were inadequate in their upper range for complete information as to the ability of pupils in the 4B grade and seemed to require the addition of a few problems of greater difficulty. Some of the types were unnecessary and could be omitted. The tests were therefore revised as has been described in the foregoing chapter.

(b) It seemed to be inadvisable to have all teachers familiar with both methods of instruction, or to have both methods used in any one school if that could be avoided.

(c) The factor of supervision was not measured and it was impossible to tell the extent to which the work of the experiment had been supervised.

(d) The measurement of actual improvement in problem solving ability as determined by the increase in number of problems solved (reasoned) correctly, regardless of the point at which improvement began, was probably an unsatisfactory criterion for judging the relative efficiency of the two methods of instruction employed. In a progressive difficulty problem test an improvement of 1.5 problems at the lower end of the "scale" may not be at all equal to a like improvement at the upper end. This might be controlled by a more accurate construction of the tests to make them real scales and by giving a "weight" to each problem, but it was considered more advisable to have all classes in the two experimental groups start at the same average level. This was done in the Reading experiment, and the manner of selecting the classes to be used is described later. The two groups were equalized with respect to their problem solving ability.

As in Philadelphia, the experiment in Reading was conducted with pupils of grades 3B, 3A, 4B, and 4A. In Reading the A sections are the higher groups. All of the pupils of the city in those grades were given the preliminary test. The test used was the revised form shown on page 00 of the Appendix. Each pupil was given a mimeographed copy of the test, and the teachers were supplied with the Test Directions used in the previous experiment. The Class Record Sheet differs slightly from that used in the Philadelphia experiments and is shown on the following page.

2. Selecting the Experimental Classes

As a preliminary step in the experimental program, the first test was given to all of the pupils in grades three and four in the city of Reading. Score sheets were obtained from the teachers and these gave the average number of problems reasoned correctly by each class. In a conference with the grade supervisors and the Director of the Reading Bureau of Research,

CLASS RECORD SHEET

Grade.....School.....Teacher.....

PROBLEM NUMBER	TALLY SHEET—NUMBER CORRECTLY REASONING PROBLEMS			
	PRELIMINARY	TOTAL	FINAL	TOTAL
1.....				
2.....				
3.....				
4.....				
5.....				
6.....				
7.....				
8.....				
9.....				
10.....				
11.....				
12.....				

A—Total.....	A—Total.....
B—No. in Class.....	B—No. in Class.....
C—Average	C—Average
A ÷ B.....	A ÷ B.....

the classes to be included in the experiment were selected by the following procedure:

1. Classes of the same, or approximately the same average scores were distributed into two groups.
2. Rejection was made of classes
   (a) With abnormal scores—high or low.
   (b) Having certain teachers who for various reasons were withheld from the experimental work.
3. The two groups of remaining classes were further examined and redistributed to
   (a) Equalize composite average scores.
   (b) Obtain classes in different and comparable sections of the city.
   (c) Maintain as nearly as possible (in the judgment of the grade supervisors) equal ability of teachers.
   (d) Avoid having two classes of the same grade in any one school taught by different methods.

As a result of this selection, six groups of classes were obtained for the experiment, each pair of classes starting at about the same level of problem solving ability as determined by the

preliminary test. The following is the line-up of the classes at the start of the experiment:

29 classes	29 classes
28* teachers	29 teachers

TABLE 5

Distribution of Pupils and Class Scores, by Groups, in Number of Problems

GRADE	NUMBER OF CONTROL PUPILS	CONTROL GROUP AVERAGE	NUMBER OF EXPERIMENTAL PUPILS	EXPERIMENTAL GROUP AVERAGE
3B.....	130	2.10	145	2.10
3A.....	172	3.67	180	3.56
3B–A.....	224	3.81	221	3.76
4B.....	95	5.12	108	5.18
4A.....	221	7.00	215	7.00
4B–A.....	191	5.07	196	5.02
Total...	1033	4.40	1065	4.25

As in the first Philadelphia experiment, the purpose of this study was again to determine the relative efficiency of two methods of improving problem solving ability. One group of classes again used the experimental method described on page 61 of the Appendix. The other group used a control method which is submitted here. This approaches a "no-method" type of instruction. To secure uniformity, however, it was necessary to set a definite line of procedure. All the teachers in the experiment were paired as equally as possible according to their teaching ability as recognized by the supervisors of the grades, to equalize the teaching strength in the two groups and at the same time maintain equality of class ability as evidenced by average scores in the first test. In no case were parallel grade teachers using different methods in the same school. All the teachers were met in conference in separate groups according to the method to be used, and the method to be used was explained by them by the experimenter. The principals and

* In one school one teacher taught two 4A classes—departmental plan.

the grade supervisors supervised the instruction throughout the seven weeks of experimentation to assure adherence to the prescribed methods.

3. The Control Method in Reading, Pa.

With the Control method there will be no definite method of *instruction* in problem solving except as noted below. No attempt at analysis or an analytic method is to be made, and teachers will please avoid the use of any scheme or plan for teaching pupils to solve problems to conflict with the prescribed plan of this method.

1. The problems to be used may be presented on the blackboard, on mimeographed or printed sheets, or selected from the pupils' arithmetic or problem text books. In the 3A grade the problems may be read aloud to the class from the blackboard, sheets, or book, without comment of any kind by the teacher.

2. At the first problem lesson of each week the teacher will spend the first half of the period in consideration of the first few problems in the manner indicated below, as many as the time permits. She will clear up language difficulties, if they exist, help pupils to an understanding of the problem *situation* involved, and obtain from the pupils the decision as to *how* the problem is to be solved. The teacher will then do the necessary work on the blackboard to obtain the correct answer and write the answer with its appropriate "label".

3. The pupils are then to be given paper and directed to find the answer to as many problems as they can in the time remaining. The teacher should prepare enough problems to keep every member of the class busy during the entire written period. The essence of this method is that pupils should learn to solve problems by actually working upon a large number of problems, with a minimum amount of direct aid from the teacher.

4. During the written period the teacher will give no *class* aid but will answer questions of individuals and give specific help on the problem presenting difficulty. She will go around her class to keep every one working and to help the individuals who are unable to find a solution. The second problem lesson of each week is to be spent entirely in this manner.

Illustration

Problem: "How many apples will Tom need to fill 4 baskets if he puts 6 apples into each basket?"

1. James, read the problem carefully.
2. What is Tom to do? How many baskets are there? How many apples is he to put into each basket? etc.
3. What are we asked to find?
4. Shall we add, subtract, multiply, or divide?

The teacher then does the necessary computation on the blackboard. She then asks, "What is the name of the answer?" The answer is then written.

4
6
—
24 Answer, 24 apples.

Illustration

Two-step Problem: "Fred worked for 3 hours one afternoon and for 6 hours on Saturday. If his pay was $.25 an hour, how much did he earn in all?"

1. Harry, read the problem carefully.
2. What did Fred do? On how many days did he work? How long did he work on these days? How much did he receive for each hour's work?
3. What are we asked to find?
4. Before we can find how much he earned, what must we know first?
5. How shall we find how much he earned? What must we know first?
6. What must we find next?
7. Shall we add, subtract, multiply, or divide?

The teacher does the necessary computation on the blackboard and then asks, "What is the name of the answer?" The answer is then written.

3 $.26
6 9
— ———
9 $2.25 Answer, $2.25

Procedure

Lessons twice a week, any days.

Duration of lessons: Grades 3A and 3B, 30 minutes.
Grades 4A and 4B, 35 minutes.

First weekly lesson—half period spent in class work, with teacher.

Second weekly lesson—entire period spent in written work by pupils.

As far as possible the class work should be marked each day and returned to the pupils. Problems that presented general class difficulty may be considered in manner suggested. General class lessons may be marked for correctness of answer. In the two experiment tests, however, the problems should be marked only on the basis of correct reasoning. See test directions.

4. Teacher Training and Experience

A study of the teachers included in the experiment showed that the teachers of the control group as a whole, and in each grade, had more teaching experience. This group also included four teachers of experience exceeding 40 years who were not graduates of a normal school. The range of experience of the teachers in this control group was from 2 to 49 years, with an average teaching experience of 15.5 years.

The teachers of the experimental group show a range in teaching experience from 1 to 37 years, with an average of 11.3 years. All of the teachers in this experimental group had received normal school training.

The following table indicates the teacher experience by grades:

TABLE 6
EXPERIENCE OF TEACHERS, BY GRADES, IN YEARS

GRADE	EXPERIMENTAL GROUP	CONTROL GROUP
4A	12.0	16.4
4B	7.8	18.4
4A–B	4.6	8.2
3A	18.5	20.0
3B	3.2	7.0
3A–B	15.2	22.1
Total	11.3	15.5
Range	1 to 37	2 to 49

5. Supervision

Only three of the elementary schools of Reading are under the direction of supervising principals. There are two supervisors of grades for the city; one for grades 4-6, and one for Kgn.-3, and it was under their active and sympathetic supervision that the work was carried on. In order that there might be a definite record of their direction, as far as it was possible to obtain such evidence, the following general letter was sent to the supervisors and the three principals:

To the Supervisor and Principal:

In all experimental work the matter of supervision is an important factor. To be sure of valid results, the necessary supervision should guarantee adherence to number and length of lessons, to correct testing procedure, and to prescribed method. Scientific approval depends on confidence in the supervisory program.

At the end of the experiment, will you please indicate on this sheet the extent of the supervision you have been able to give to this experiment by stating the number of times your definite supervisory activity covered these factors:

1. The number of visits made to classrooms to see problem lessons taught in (a) Control classes.....
   (b) Experimental classes.....
2. The number of conferences or meetings held with teachers in the experiment in groups.....
3. The number of individual conferences with teachers on some phase of the problem experiment.....
4. The number of times you have given personal assistance in this work to individual teachers, or have made corrections or adjustments in methods of teaching
   (a) in demonstration work.....
   (b) in correction or adjustment of method.....
   (c) in specific help at request of teacher.....

Very truly yours

Roy E. Adams.

Signed

.....

Supervisor-Principal

There is evidence that the experiment was carefully supervised, both by the supervisors and the principals. The reports made in reply to this letter are tabulated on the following page.

The grade supervisors made one visit to each of the 57 teachers in the experiment, and in the schools that had supervising principals (three in number), 44 visits were made. It is noteworthy that 16 minor corrections or adjustments in method were made and a total of 60 individual conferences with the teachers about the work. That the three principals held 6 group meetings and 17 individual conferences is a pleasing indication of the careful supervision the work received in those schools. Assurance was received from the supervisors that each group received

TABLE 7

SUMMARY OF THE REPORTS ON SUPERVISION

	SUPERVISORS	PRINCIPALS	TOTAL
Visits to classrooms to observe lessons			
a. Control classes	23	15	38
b. Experimental classes	24	29	53
Conferences or group meetings	8	6	14
Individual conferences	43	17	60
Personal assistance			
a. Demonstration	0	2	2
b. Correction or adjustment	11	5	16
c. Special help teacher requested	2	2	4

an equal amount of attention and that the individual conferences were almost equally divided between the two groups.

The final test, which has been described in Chapter 2, with the method of its formulation, was administered at the end of the experimental period to determine the extent of the progress made. In the same manner as before, class *averages* were determined and comparisons were then made with the scores of the preliminary test. From the differences in number of problems correctly reasoned, as indicated by the class averages for each group, it was possible to draw certain conclusions as to the relative value of the two methods of instruction used in the experiment.

6. Comparison of Results

The summary of the results of the two tests is shown in the tables on the following pages, with the statement of the average gain made by the grade group in the number of problems correctly reasoned.

Grade 3B (low third). In this grade the gains range from .94 to 2.17 in the experimental group, with an average gain of 1.41. The control group has a range in gains of from .50 to 2.34, for an average of 2.07. In this group the Control method shows a greater gain by .61.

Grade 3B-A ("double" grade). With the exception of one class, the gains in the experimental classes of this grade group

were practically uniform, for an average of 1.83. The control group, however, shows less average gain and quite a wide divergence in individual class gains.

Grade 3A (high third). In spite of the large gain of 3.10 made by one of the control group classes, the experimental group shows a greater average gain: 1.96 as compared with 1.78. The gains in the experimental group were fairly uniform.

Grade 4B (low fourth). The control classes of this group made a consistent gain of about 2.5 problems while the experimental group shows a gain of but 1.27 problems.

Grade 4B-A ("double" grade). Both groups gained to about the same extent in this grade, a slightly greater increase being made by the control group: 2.48 to 2.23.

Grade 4A (high fourth). The gains in this grade were about the same, with a slightly greater increase in the control group. It must be noted, however, that the average gain of the control group was affected greatly by the two classes taught by the same teacher, gains of 5.39 and 3.11 being made, one of these gains being greater that that made by any other class. The Supervisor in charge states that this may have been due to (1) high quality of teaching and (2) grouping by ability which placed in these classes only pupils of higher I. Q. To that extent, the results in this grade may not be as significant as they appear on the surface.

Summary of Results

A consideration of the results in all the grades shows no consistent difference in gains made, regardless of the method used. To claim a general superiority for either method is not supported by the evidence secured. It is to be noted, however, that the analysis method produced greater gains in two of the third grades, a result similar to the one to be seen in the final experiment in Philadelphia. The large gain made by the control group in grade 4B has been discussed.

Summary of Conclusions

On the basis of the results obtained from seven weeks of experimentation in the Reading schools we may make the following general conclusions:

1. Pupils in the lower grades, who are just beginning the work in problem solving, seem to be helped by a training in detailed analysis of the problem presented.

2. In the fourth grades, the method that emphasizes the solving of many problems without specific class instruction in method of attack produced better results.

3. These conclusions, on the whole, seem to be in harmony with the results of previous studies with pupils in higher grades.[1]

TABLE 8

SUMMARY OF THE RESULTS IN THE TWO TESTS

GRADE	TEST 1		TEST 2		GAIN		EXCESS GAIN
	EXP.	CONT.	EXP.	CONT.	EXP.	CONT.	
4A.....	7.00	7.00	9.00	9.20	2.00	2.20	Cont.— .20
4A–B.....	5.02	5.07	7.25	7.55	2.23	2.48	Cont.— .25
4B.....	5.18	5.12	6.45	7.61	1.27	2.49	Cont.—1.22
3A.....	3.56	3.59	5.52	5.37	1.96	1.78	Exp.— .18
3A–B.....	3.74	3.81	5.57	4.81	1.83	1.00	Exp.— .83
3B.....	2.10	2.10	3.51	4.17	1.41	2.07	Cont.— .66

[1] Washburne, C. W. and Osborne, R., "Solving Arithmetic Problems", *Elementary School Journal*, Dec., 1926, pp. 296-304.

IV

The Philadelphia Experiment, Oct. to Dec., 1929

A. Schools and Classes in the Experiment

Fifteen Philadelphia public schools were selected for the experimental work. They were schools considered to be representative of conditions in a large city and included:

a. Different social and economic conditions. These schools were from (a) good neighborhood sections, providing pupils from average well-to-do families, (b) industrial sections, with children of laborers and artisans, and (c) the so-called "poor" sections of the city.

b. Many racial and national elements. (a) The pupils of some of the schools were largely of Jewish and better class American parents. (b) Other schools were composed of the many nationalities usually found in industrial sections. (c) The other schools included a large percentage of Negro pupils and children of Italian parents.

c. Different sections of the city.

d. Schools in which the Principals were willing to engage in the experimental work. All of the Principals were informed of the nature of the work and the character of the supervision that they would be expected to give to it.

e. Schools at about the same general level of intelligence as indicated by results in the Philadelphia Group Test of Mental Ability given to 6B grades.

f. Schools at approximately the same level in problem solving ability as determined by a test given by the Philadelphia Division of Research in March, 1929.

The distribution of these 15 schools into the various economical, national, and location groups is shown in the following table:

TABLE 9

DISTRIBUTION OF SCHOOLS IN PHILADELPHIA

A. By social and economic conditions
   1. Good, residential ..... 4 schools
   2. Industrial ..... 5 schools
   3. Poor, lower class ..... 6 schools
B. By race and nationality
   1. Jewish and American parents ..... 4 schools
   2. Negro and Italian parents ..... 6 schools
   3. Numerous nationalities ..... 5 schools
C. By city location
   1. North Philadelphia ..... 5 schools
   2. South Philadelphia ..... 6 schools
   3. West Philadelphia ..... 4 schools

These fifteen schools finally chosen were of practically the same relative standing in intelligence, general problem solving and reading ability. There were seven schools in the group that later used both methods of instruction during the experimental period, four schools that used only the experimental method, and four schools that used only the control method.

During November, 1929, the Philadelphia Group Test of Mental Ability was administered to the pupils of the 6B grades in nearly all Philadelphia schools by the Philadelphia Division of Research. The scores[1] of the schools in this experimental study, grouped into three classes according to the type of method used during the experiment, show the following average intelligence scores and the average mental ages derived therefrom:

TABLE 10

AVERAGE SCORES ON PHILADELPHIA GROUP TEST OF MENTAL ABILITY IN 6B GRADES OF 15 SCHOOLS IN THE EXPERIMENTAL PROGRAM, WITH DERIVED AVERAGE MENTAL AGES

	AVERAGE OF GRADE 6B INTELLIGENCE SCORES	AVERAGE MENTAL AGE
Experimental—4 schools	48.8	11 yrs. 10 mo.
Control—4 schools	47.5	11 yrs. 10 mo.
Both—7 schools	47.8	11 yrs. 10 mo.

[1] Test results supplied by courtesy of the Philadelphia Division of Research.

This table clearly indicates that the schools were on an equal level in the factor of intelligence as determined by a comparison of their 6B grades. It is not a statement of equality of intelligence in the grades 3A-4B included in the experiment but it is probably indicative of an equality that may be safely assumed to exist in all of the grades in those schools. Unless there had been a great change in pupil make-up of these schools and in character of the neighborhood, both of which conditions certainly were not evident, the intelligence level of the 6B grades should be representative of the schools as a whole.

In order to determine further the comparative standing of these schools, a study was made of the results of a problem solving test administered in March, 1929, as a part of a survey testing program by the Philadelphia Division of Research. The Division very kindly supplied the data for the 15 schools, and the following table shows the average school scores for grades 3B to 5A inclusive. As this test was administered early in the term, it could not be given to 3A pupils. The scores of the 5A pupils are included as indicative of 4B pupils who have recently left that grade. These scores therefore represent the problem solving ability of pupils of grades 3A to 4B in the schools included in the experiment.

TABLE 11
SCHOOL AVERAGE SCORES IN SURVEY PROBLEM SOLVING TEST GRADES 3B–5A, MARCH, 1929, IN PHILADELPHIA, PA.

	NUMBER OF SCHOOLS	SCHOOL AVERAGE SCORE
1. Experimental	4	4.90*
2. Control	4	4.36
3. Both Methods	7	4.41
Total	15	Combined—4.53

This table indicates a fairly equal statement of ability of schools on the basis of problem solving. This was supplemented by the experimental procedure which selected the classes from

* This score is raised by the record of one 3B class of one school. That class made a score very much higher than any other 3B class in the other 14 schools. The experimental work did not include this class.

these schools that presented, by pairs, almost equal ability in problem solving as shown by the results in the preliminary test of the study. This selection is explained in detail on page 00.

The relation of reading ability to intelligence is generally recognized as particularly significant. By many, a good reading test is considered a close equivalent to an intelligence test for purposes of pupil classification by groups, and pupils grouped on the results of reading tests may be said in general to be grouped by intelligence. It is also recognized that reading is an important factor in problem solving ability as shown in studies by Wilson,[2] Greene,[3] Lessinger,[4] and Boyer.[5]

The schools concerned in this experimental work were studied to determine their reading ability as evidenced by records of the Philadelphia Division of Research in the Stanford Achievement Test in Reading, Paragraph Meaning, Form B administered as a part of the survey testing program in September of the experimental year. Median results are presented in terms of grade unit scores for eleven of the schools participating in this experimental program. The *average* grade unit scores are here given for the three groups of schools already indicated: experimental, control and both methods:

TABLE 12

STANFORD ACHIEVEMENT TEST IN READING—PARAGRAPH MEANING, FORM B
AVERAGE MEDIAN GRADE—UNIT SCORES BY GRADE SECTIONS FOR 11 SCHOOLS IN THE PROBLEM SOLVING EXPERIMENT

	3A	3B	4A	4B
Experimental (3 schools)	3.0	3.7	4.5	4.9
Control (4 schools)	3.0	3.6	4.3	4.7
Both (4 schools)	3.2	3.8	4.7	5.4
TOTAL (11 schools)	3.1	3.7	4.5	4.9

[2] Wilson, Estaline, "Improving the Ability to Read Arithmetic Problems", *Elementary School Journal*, XXII, pp. 380-86.

[3] Greene, Harry A., "A Directed Drill in the Comprehension of Verbal Problems in Arithmetic", *Journal of Educational Research*, X (Nov. 1925), pp. 287-300.

[4] Lessenger, W. S., "Motivation and the Accomplishment Quotient Technique", College of Education Library, University of Iowa, 1925. "Reading Difficulties in Arithmetic Computation", *Journal of Educational Research*, XI, 1925, pp. 287-291.

[5] Boyer, Philip A., "Diagnosis in Problem Solving in Arithmetic". Sixteenth Annual Schoolmen's Week Proceedings, U. of P., 1929, pp. 357-364.

These results, in the form of averages of median scores for the grades, are obtained from mimeographed reports of the Philadelphia Division of Research, and they show that the schools were on a practically even level in reading ability. The larger scores apparent in the "both method" schools do not affect this conclusion as the two teaching methods were used in those schools.

We may summarize our consideration of the equality of the schools included in the experiment by stating briefly that they were:

1. Representative of conditions in a large city,
2. On an equal level in the factor of intelligence as determined by a study of their 6B grades,
3. Equal in problem solving ability as indicated in the results in tests administered by the Philadelphia Division of Research in March, 1929, and,
4. Equal in reading ability as indicated by results in the Stanford Achievement Test in Reading—Paragraph Meaning, Form B.

A further study of the equality of the groups included in the experiment concerned itself with teacher equivalence. This is discussed later. Evidence was secured that indicated approximate equality in training, experience, and post-normal school professional training.

Selection of Classes

The first test, which was previously discussed in Chapter 2 and is shown of page 59 of the Appendix, was administered to the 117 classes in the 15 schools selected, and was the basis for the grouping of the classes for the experimental work. Class median scores were obtained and were used for "pairing" the classes. Classes were paired when their medians were equal or almost equal.

When it was possible, the two methods of instruction were not used in the same school, but there was no instance of the two methods used in the same numerical grade in the same school. The teachers of all the classes in the experiment were instructed to avoid comparison of methods while the work of the experiment was in progress and to refrain from any discussion of their work with other teachers.

As far as was possible the various sections of the city were equally represented. No double grades were selected, viz. 3A and 3B combination classes. In those schools having grouping by ability care was used not to have different ability groups competing in the experiment even though their median scores in the first test were the same.

The following served as a basis for the class selection:

1. Equal or nearly equal class median scores.
2. The two methods not used in the same school, or not used in the same numerical grade in any school.
3. Classes with approximately equal number of pupils.
4. Different "ability groups" not paired.
5. Different sections of the city represented.
6. No "double grades" selected.

Of the 117 classes that took the first test it was necessary to exclude 21 from the experiment because (a) they were double grades, (b) their medians were too high or too low and were not equalled by any other classes, (c) there was no class with a median sufficiently equal, or (d) the number of pupils in the class was too large or too small.

The 96 classes finally selected for the experiment are shown comparatively in the following table:

TABLE 13
COMPARISON OF THE 96 CLASSES AT THE BEGINNING OF THE EXPERIMENT

	EXPERIMENTAL	CONTROL
Total number of pupils	1938	1836
Number of pupils by grades—3A	574	531
3B	365	362
4A	565	522
4B	434	421
Average size of classes—3A	41.0	36.9
3B	40.5	40.2
4A	40.3	38.4
4B	39.4	38.2
Number of classes included	48	48
Average of class median scores—3A	*2.79*	*2.83*
3B	*4.01*	*4.00*
4A	*5.74*	*5.60*
4B	*7.34*	*7.31*

It is evident that these two groups, grade by grade, were at approximately the same level in problem solving ability at the same time the experimental work began. The relative growth of these groups during the experimental period should indicate the relative values of the two methods of instruction during that period if other factors remain constant or controlled.

Validity of Selection. In dealing with large numbers of pupils in any experimental work it is probably justifiable to assume that many variables will be neutralized by the mere fact of numbers. The groups selected include nearly 2000 pupils in each, they are taught by 96 teachers about equally divided as to training and experience, and they are composed of pupils from 15 different schools selected as particularly representative of the social, economic, and racial conditions obtaining in a large city. It may probably be safely assumed, also, that in groups of this size the average intelligence level would be about the same. These pupils have received problem instruction for periods ranging from one year to two and a half years in their regular class routine. Differences in intelligence would have produced differences in problem solving scores had they existed. It is not likely that eight additional weeks of instruction during the experimental period would produce significant differences in results that had not made themselves evident in the preceding class life, or differences that could be claimed as due to different intelligence capabilities. Evidence has been presented to show that these pupils, by school scores, were of approximately the same reading ability, problem solving ability, and intelligence rating. The teachers of the two groups have also been shown to be approximately equal in training, experience, and post-normal school professional training.

It will be remembered that all classes with very high or very low median scores were excluded from the experiment, and that in no case was a "high ability" group paired with a "low ability" group.

Intelligence and Problem Solving Ability

Of the results secured another phase which is not the primary purpose of this study, but which is included for the value it

may have to students of arithmetic and for its value as a check on the final selection of the experimental classes, concerns itself with the correlation between intelligence (I. Q. scores) and problem solving ability as evidenced by scores in various arithmetic reasoning elements in standard intelligence tests and by scores obtained from the test used in this study.

When classes are paired according to equality in problem solving ability in large numbers, it is probably true that they would also be paired equally in intelligence possibilities since there exists a high correlation between the factor commonly called "intelligence" and the ability called "reasoning" in arithmetic. The following studies are cited:

1. The correlation between the I. Q. scores obtained from the Otis Classification Test, Part 11, and the scores in arithmetic reasoning obtained from the Otis Arithmetic Reasoning Test (Test 5 of Otis Group Intelligence Scale—Adv. Exam.), a study that included 196 cases of 8B pupils in a Philadelphia Junior High School. The writer converted pupils' scores into arithmetic ages by means of the conversion table of the test. Arithmetic quotients were then derived by dividing the arithmetic ages by the chronological ages, both expressed in months. These arithmetic quotients were then correlated with the intelligence quotients previously determined, and by means of the Otis Correlation Chart (product-moment formula) a coefficient of correlation of .638 was obtained. The table of A. Q. and I. Q. scores follows:

TABLE 14

CORRELATION BETWEEN ARITHMETIC QUOTIENTS AND INTELLIGENCE QUOTIENTS FOR 196 PUPILS OF 8B GRADE IN A JUNIOR HIGH SCHOOL OF PHILADELPHIA

$r = .638$

INTELLIGENCE QUOTIENTS	ARITHMETIC QUOTIENTS								
	57–60	61–70	71–80	81–90	91–100	101–110	111–120	121–130	131–140
121–130								1	
111–120				1	2	2	4	1	1
101–110			3	5	14	23	5	3	1
91–100		4	20	15	16	6	7		
81–90.	5	8	11	11	6	2			
71–80.	1	8	7	1					
61–70.		2							
51–60.									

2. The correlation[6] between I. Q. scores obtained from the Trabue Completion Test, B and C, and the arithmetic reasoning element of the Army Alpha Test, Part IV, a study that included 287 pupils of grades 4-8, showed a correlation coefficient of .65.

3. The correlation was determined by the writer between the scores obtained from 90 pupils of grade 4A on the Haggerty Intelligence Test and the arithmetic problems of Test 1 (see Appendix, page 59) used in the problem solving experiment in Philadelphia. For this correlation, the individual scores of 90 pupils in the 4B grade of one of the schools of Philadelphia engaged in the experimental work were obtained. To these same pupils the Haggerty Intelligence Test was given, and the scores in that test were correlated with the problem solving scores made by these pupils in Test 1 of the experimental program. The Otis Correlation Chart was used to determine the coefficient of correlation and gave a result of .548 ± .049.

TABLE 15

CORRELATION OF SCORES IN THE HAGGERTY INTELLIGENCE TEST AND TEST 1 OF THE EXPERIMENTAL PROGRAM; FOR 90 PUPILS OF GRADE 4A IN ONE PHILADELPHIA SCHOOL

SCORES IN HAGGERTY TEST											
91–100											
81–90								1	1		
71–80								2			
61–70					1			1		4	
51–60			2		1	1	2	4	1	2	
41–50		1	1	1	4	4	11	7	2	3	
31–40		2	1	3	3	3	3	6			
21–30			3	2	2				1		
11–20		1	1			1		1			
0–10											
Scores in Test 1	0	1	2	3	4	5	6	7	8	9	10

r = 548 ± .049

4. The writer made a study of 606 elementary school pupils in one Philadelphia school to discover the correlation between

[6] Memoirs of the National Academy of Sciences, Vol. XV, "Psychological Examining in the United States Army", p. 337. Washington Government Printing Office, 1921.

the I. Q. scores obtained from the National Intelligence Test and arithmetic quotients derived from the arithmetic reasoning element in that test. The arithmetic scores were converted into arithmetic age scores, and from those were obtained arithmetic quotients. These arithmetic quotients were then correlated with the intelligence quotients derived from the complete test. The coefficient of correlation (Otis Correlation Chart used) was found to be .734 ± .012. It must be noted that the arithmetic reasoning score appeared twice in this correlation, in the total score from which the intelligence quotients were obtained and as a separate factor in determining the arithmetic quotients. This probably increased the coefficient of correlation somewhat, but the result is still probably significant enough to justify the claim that intelligence and problem solving ability are highly correlated.

These four studies clearly indicate that a high correlation exists between problem solving ability and intelligence. As the classes in this experiment were paired on the basis of their problem solving ability, it is therefore safe to assume that they were also paired on the basis of intelligence.

B. Experimental Data

The experiment included grades 3A, 3B, 4A, and 4B in fifteen representative public schools of Philadelphia, under the guidance of 96 teachers. It involved a total of 3774 pupils. It began with a preliminary test on October 24, 1929 and ended with the final test on December 20, 1929. There were two problem solving lessons a week in each grade, in accordance with detailed instructions in method as shown on pages 00 of the Appendix. The length of lessons was set at 30 minutes in grades 3A and 3B, and at 35 minutes in grades 4A and 4B. In Philadelphia the B section is the higher.

Each teacher in the experiment received copies of the following sheets of information:

a. TEST DIRECTIONS—See Appendix, page 60
b. CLASS RECORD SHEET
c. TEACHING METHODS, Experimental or Control See Appendix, page 61-64

TABLE 16

CORRELATION OF INTELLIGENCE AND ARITHMETIC REASONING FOR 606 ELEMENTARY SCHOOL PUPILS IN ONE PHILADELPHIA PUBLIC SCHOOL

INTELLIGENCE QUOTIENTS FROM—TO	ARITHMETIC QUOTIENTS FROM 54 TO 58.9	FROM 59 TO 63.9	FROM 64 TO 68.9	FROM 69 TO 73.9	FROM 74 TO 78.9	FROM 79 TO 83.9	FROM 84 TO 88.9	FROM 89 TO 93.9	FROM 94 TO 98.9	FROM 99 TO 103.9	FROM 104 TO 108.9	FROM 109 TO 113.9	FROM 114 TO 118.9	FROM 119 TO 123.9	FROM 124 TO 128.9	FROM 129 TO 133.9
134–138.9															1	
129–133.9														1	1	
124–128.9									1	2			1	2	2	
119–123.9						1					2	2	2		1	
114–118.9									2	5	8	2	6	1	2	
109–113.9							2	3	3	3	5	8	4	2		1
104–108.9				1	1	1		6	8	9	10	9	3		1	
99–103.9				1	1		9	11	14	13	7	3	3	1		
94– 98.9						5	6	18	15	6	3	2	1			
89– 93.9				1		6	12	19	14	8	6	1				
84– 88.9		1			3	16	24	23	9	6	4	1				
79– 83.9			1	7	12	14	11	9	9	1	2	1				
74– 78.9			1	8	10	18	12	7		1						
69– 73.9				2	13	3	11	4	2	2		1				
64– 68.9		1	2	6	4	3	5	1								
59– 63.9		1		4	3	1	3									
54– 58.9	1	1	1		2		1									
49– 53.9			1		1		1									
44– 48.9	2						1									

$r = .734 \pm .012$

The Test Directions gave detailed information as to the manner of administering the tests, the scoring of the papers, and the tabulating of results. A mimeographed copy of the test was given to each pupil and he solved the problems on a separate sheet of paper. The test was marked on the basis of correct reasoning only (see Test Directions) and the results tabulated as class median scores on the record sheet. These scores are representative of the pupils' reasoning ability and do not take accuracy into account. They were used as the basis for the selection of the final experimental classes as has been previously explained. The CLASS RECORD SHEET is shown below. A similar sheet was used at the final test in December, and from the two records information was obtained as to comparative results of the methods.

CLASS RECORD SHEET

Grade..... Teacher.....

School.....

PUPIL SCORE	NO. OF PUPILS
12	
11	
10	
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	

No. of Pupils.....
Class Median.....

Date.....

Just as in the two preliminary studies the purpose of this final Philadelphia experiment was to determine the relative

efficiency of two differing methods of improving problem solving ability. Half of the classes finally selected were taught in accordance with the experimental plan described in detail on page 61 of the Appendix, and were known as the Experimental Classes. They used the method which is analytical to considerable detail. The other classes were taught by the control method, the method of teaching problem solving prescribed by the Philadelphia Course of Study in Arithmetic. This method is shown in detail on page 63 of the Appendix.

In no case were teachers of parallel grades using different methods in the same school. All teachers were given copies of the method to be used, and the principals of the schools were informed as to the particular procedure that their teachers were to follow. It will be noted that the method instructions included an outline of the teaching steps involved, illustrations of the methods as used with a typical problem, and a statement of the procedure to be followed with respect to specified number, length, and nature of problem lessons during the experimental period.

C. The two methods differ in these essential points: (a) the *amount* of analysis involved in solving the problems, and (b) the procedure during class instruction. The experimental method is an *extension* of the control method and includes several additional steps in the detail of analysis and reasoning. It also provides the opportunity for class consideration of more problems during a lesson period, in that it emphasizes the necesssity for making a choice of operation of many problems rather than the complete solution of a few problems.

D. Teacher Training and Experience

At the beginning of the experiment, information was secured as to the training of each of the 96 teachers doing the experimental work, the number of years of teaching experience, and the number of semester hours of post-normal school training. The following letter was sent to each of the Principals:

November 11, 1929.

In connection with the experiment in problem solving I would like to have a record of the training and experience of the teachers involved. Will you

please have the information placed on this letter, at the bottom, as indicated, for the teachers in your school. There should be no hesitancy in giving this information as no names of teachers will appear in the report.

Very many thanks for your help.

Sincerely yours,
Roy E. Adams.

TEACHER	NORMAL SCHOOL GRADUATE YES OR NO	YEARS OF EXPERIENCE	NUMBER OF SEMESTER HOURS OF POST-NORMAL TRAINING

The information was secured for 83 of the 96 teachers involved and shows that the experimental and control groups were of practically equal standing in these factors. One teacher in each of the groups was not a normal school graduate. The others had received the regulation two-year normal school training. The average number of years of experience for the teachers in the experimental group was 11.75, with a range from 2 years to 46 years. The teachers of the control group averaged 12.15 years in experience, with a range from 2 years to 40 years. The amount of post-normal school training received by the teachers of the two groups averaged about the same, 9.16 semester hours for the experimental group and 8.45 semester hours for the control group.

It was not possible to determine the comparative teaching ability of the teachers engaged in the work. They were in 15 different schools and were rated as to ability by as many different principals. Such ratings would be of little comparative value. As comparative ratings by one individual or by any single group of individuals were not available, it was necessarily assumed in this study that the large number of teachers involved, 48 in each group, would guarantee approximately equal teaching

ability. The study has produced evidence to show that these teachers were about equal in training, experience, and after-school professional training.

TABLE 17

AVERAGE TRAINING AND EXPERIENCE OF TEACHERS

	EXPERIMENTAL	CONTROL
Normal school graduate?	Yes, 35; No, 1	Yes, 46; No, 1
Years of experience—Average	11.75	12.15
Range	2 to 46 yrs.	2 to 40 yrs.
Post-normal school training in semester hours—Average	9.16	8.45
Range	0 to A.B. degree	0 to 48 hrs.
Number of teachers reported	36	47

E. Supervision

In order that the factor of supervision might be accurately determined, the Principals of the 15 schools in the experiment were asked to record on a letter-questionnaire the extent of their supervisory activity. The following letter was sent to each of the Principals:

October 31, 1929.

TO THE PRINCIPAL:

The success and reliability of an experiment in education are often dependent upon the supervision given to it. It determines adherence to the experimental program, guarantees uniformity in number and length of lessons, correct teaching and testing procedure. Scientific approval very often depends upon confidence in the supervisory program.

That the results in our problem solving experiment may be sure of reliability in this respect, will you please endeavor to give it the supervision required, and at the end of the experiment indicate on this sheet the number of times your definite supervisory activity covered the factors listed below?

I will appreciate thoroughly your care and promptness in making such a report.

Very truly yours,

ROY E. ADAMS.

School.....Principal.....

1. The number of visits to classrooms to see problem lessons taught in
   (a) Experimental classes.....
   (b) Control classes.....
2. The number of conferences or meetings held with the teachers engaged in the experiment.....
3. The number of individual conferences with teachers on some phase of the experimental work.....
4. The number of times you gave personal assistance or made corrections or adjustments in methods of teaching, by
   (a) Demonstration work.....
   (b) Specific help at the request of the teacher.....

At the end of the experimental period a report of the supervisory activity of the Principals was obtained from eleven schools, covering the work of 36 experimental classes and 31 control classes. This report is shown here in table form:

TABLE 18

SUPERVISION BY PRINCIPALS OF 11 SCHOOLS IN THE EXPERIMENT

	EXPERIMENTAL CLASSES	CONTROL CLASSES
1. Supervisory visits by the Principal	103	107
2. Conferences in groups	15	18
3. Conferences with individual teachers	35	35
4. Personal assistance by the Principal		
a. Demonstration work	9	7
b. Help requested by the teacher	16	14

It is evident from this report that there was careful follow-up work on the part of the Principals of these 11 schools. Each group of classes received about the same amount of attention. The average number of times the teachers were visited for supervisory purposes was about 2.86 in the experimental group and 3.14 in the control group. It is significant that 33 group conferences were held in these 11 schools, an average of 3 meetings in each, indicating that more was done in this respect than a single preliminary meeting for the purpose of explaining the nature of the experimental work. The 70 individual conferences during the experimental period vary from 2 to 8 in the ex-

perimental classes and from 2 to 10 in the control group, an average of about one conference to each teacher on the number reported. The Principals further showed their interest by teaching 16 demonstration lessons, about equally divided between the experimental and control groups. The number of teachers who evidenced a desire for assistance in their work was 30, an average of about 3 out of every 7 teachers reported, and a pleasing indication of the interest they took in their work. These requests for help come from the two groups in about equal numbers.

From these returns it is probably safe to believe that the experimental program received careful and adequate supervision and that this supervisory activity was equally divided between the two groups. Assurance was received that the Principals adhered to the limits set for the experiment and that they closely followed the teaching methods assigned to their schools. As has been stated, only seven of the fifteen schools in the experimental work used both methods of instruction, and in no case were the two methods used in parallel grades in the same school. Eight schools were informed of but one method, four using the analysis method and four the control.

F. Results of the Experiment

The final tabulation of the results of the two tests shows the following growth in average class median scores in the number of problems correctly reasoned:

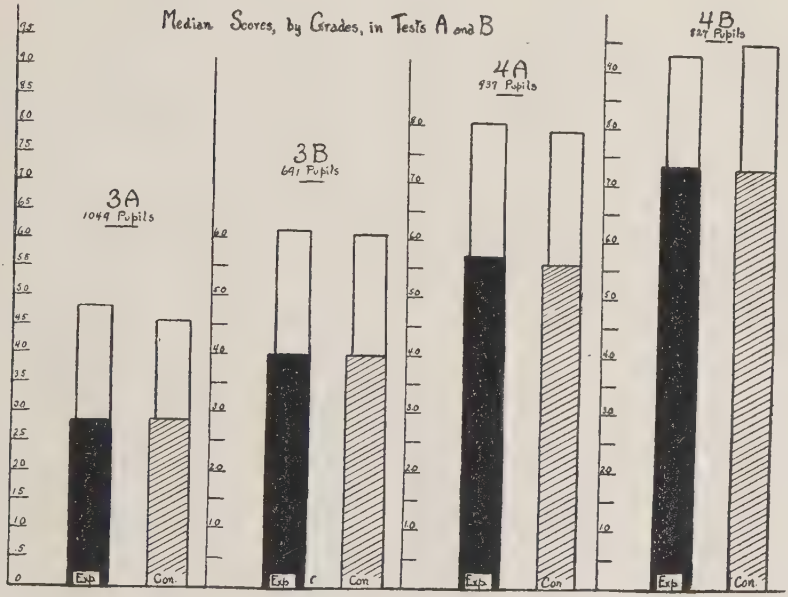
TABLE 19
AVERAGE GRADE MEDIAN SCORES IN TEST 1 AND TEST 2, WITH GAINS IN NUMBER OF PROBLEMS CORRECTLY REASONED

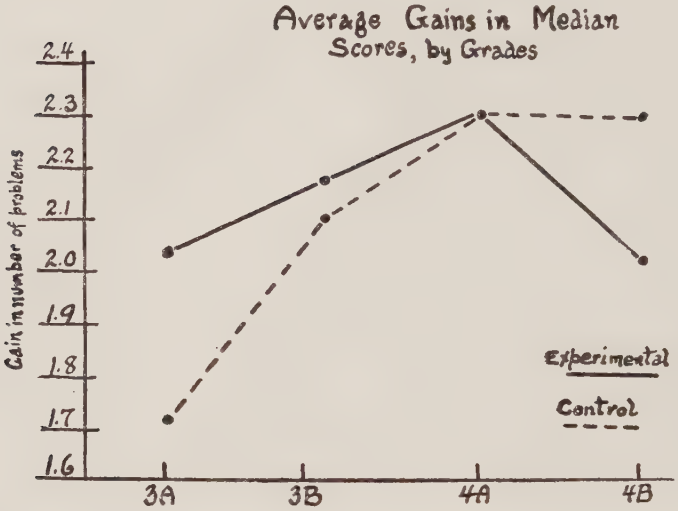
	GRADE	PRELIMINARY TEST	FINAL TEST	GAIN IN NUMBER OF PROBLEMS PER CENT
Experimental Classes....	3A	2.79	4.83	2.04–73.1
	3B	4.01	6.18	2.17–54.1
	4A	5.74	8.04	2.30–40.0
	4B	7.34	9.36	2.02–27.5
Control Classes.....	3A	2.83	4.55	1.72–60.7
	3B	4.00	6.10	2.10–52.5
	4A	5.60	7.90	2.30–40.0
	4B	7.31	9.50	2.19–29.9

TABLE 20

COMPARISON OF GAINS IN NUMBER OF PROBLEMS IN MEDIAN GRADE SCORES

GRADE	EXPERIMENTAL	CONTROL	GREATER GAIN BY	EXCESS	PER CENT
3A.....	2.04	1.72	Experimental	.32	12.4
3B.....	2.17	2.10	Experimental	.07	1.6
4A.....	2.30	2.30	.....	0	0
4B.....	2.02	2.19	Control	.17	2.4




Average Gains in Median Scores, by Grades
Gain in number of problems
2.4
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
3A
3B
4A
4B
Experimental
Control

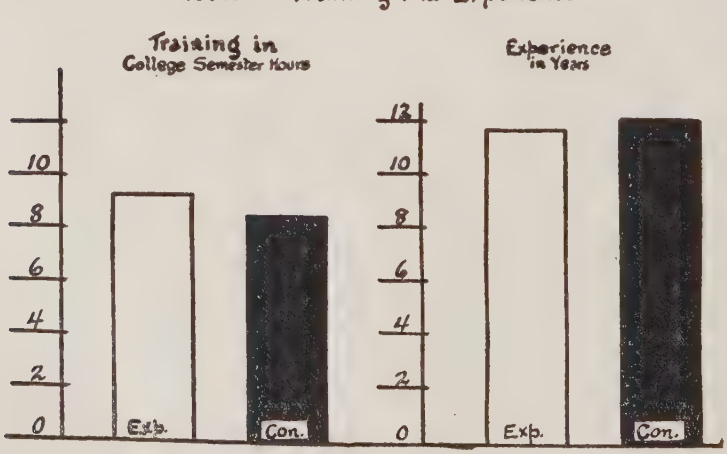

Teacher Training and Experience
Training in College Semester Hours
Experience in Years
12
10
8
6
4
2
0
Exp.
Con.
Exp.
Con.

TABLE 21

GROUP COMPARISONS, BY GRADES

	3A	3B	4A	4B
Greatest gain by	Exp.	Exp.	None	Cont.
Excess gain in number of problems	.32	.07	0	.17
Highest score of Exp. classes	7.1	7.6	10.7	11.7
(final test) Cont. classes	6.3	8.2	10.8	11.5
Greatest gain of Exp. classes	3.9	3.1	3.9	4.6
(final test) Cont. classes	2.8	3.1	4.2	2.7
Number of pupils, Exp. classes	574	365	565	434
Cont. classes	531	362	522	421

G. Statement of Results

In the foregoing tables we see the gains made by the two groups in each grade expressed in term of number of problems correctly reasoned. These gains may be expressed in terms of percentages by placing the gain for each grade over the score made in the preliminary test. When this is done, we have the results expressed in the last column of tables 22 and 23.

Grade 3A: The experimental group made the greater gain of .32 of a problem in this grade. This was an improvement 12.4 per cent greater than that made by the control group. One of the experimental classes made the highest score of the grade, and it was also an experimental class that made the greatest individual class gain in number of problems, a gain of 3.9 problems as opposed to the greatest gain of 2.8 problems in the control group.

Grade 3B: The greater gain in this grade is again made by the experimental group. The excess was .07 of a problem, an improvement 1.6 per cent greater than that made by the control group. However, a control class made the highest score of 8.2 in this grade. Neither group excelled in an individual class gain, the greatest gain in each group being 3.1 problems.

Grade 4A: The gains made by the two groups in this grade were identical: 2.3 in number of problems, or a 40.0 per cent improvement over the scores in the preliminary test. The highest score was made by a control class that increased its median from 6.6 to 10.8, a gain of 4.2 problems. The highest

experimental score was 10.7, increased from 7.6. The best individual class gain in the experimental group was 3.9 problems.

Grade 4B: In this grade the results show a greater gain by the control group, the excess being .17 of a problem, a greater improvement of 2.4 per cent. It is to be noted, however, that the highest score in the grade was made by an experimental class, 11.7 problems, and that the greatest individual class gain of 4.6 was also made by an experimental class.

Discussion of the Results

It is particularly evident in the 3A grade that the analysis method has produced a gain decidedly[7] superior to that produced by the control method. A difference of .32 of a problem, 12.4 per cent, in a short period of experimentation may have considerable weight in any claim for the superiority of the analysis method. This gain was made in a grade in which the pupils are just beginning the real work in problem solving and are not influenced by any previously taught method. These pupils began with the analysis method as their first method of solving problems. The control group of pupils began their work with the control method. Neither group had to pass through a transition stage involving first a removal of habits previously learned and an adjustment to a new method. The gain of .32 of a problem in this grade is therefore particularly significant.[8] The evidence also shows that of the 48 classes in each

---

[7] W. A. McCall in "How to Experiment in Education" says on page 163, "No matter how large the D, we can never be absolutely certain that the experimental results secured from any sampling less than the entire group are true for the entire group. Since absolute certainty is never obtainable, except for the particular group used, statisticians have coined the term *practical certainty* to designate a degree of certainty which is generally acceptable. Practical certainty is defined as plus and minus three times the SD of the measure in question." From this we conclude that the SD of the difference in gains should be plus and minus three times the observed difference in the two groups. Using the method and the formula described on pages 164 ff of McCall (*op. cit.*), we find that the SD of the differences in gains is .09. Our observed difference of .32, being more than three times that figure, is accordingly regarded as acceptable evidence of true gain made by the experimental group.

[8] The greater gain of the experimental group in the 3A grade when based on *arithmetic mean* scores as determined from the frequency distribution of the scores of all the pupils in each group amounts to .41. This figure is still more conclusive of the greater value of the analysis method in this grade.

group the highest three scores in this grade were made by experimental classes, and that five classes in this experimental group made greater individual class gains than any class in the control group.

In grade 3B the results show a gain of .07 of a problem, 1.6 per cent, in favor of the analysis method. This figure may be too small as determined by this study to justify clearly a claim for the superiority of that method. If we use the formula previously described[9] to find the SD of the difference in gains, the result is found to be .12. As the gain of the experimental group was *not* three times this SD, it would not, as far as this study is concerned, be accepted as satisfactory evidence of the superiority of the analysis method in this grade. It would be well, however, to consider the extent to which this result was affected by other conditions before a final decision is made.

In the fourth grades, so far as the external results of this study indicate there is no evidence sufficiently conclusive to justify a claim for either method. In the 4A grade, the groups show an equal amount of gain. In grade 4B, the control method has an advantage of .17 of a problem but again the highest score and the greatest individual class gains were made by experimental classes. The SD of the difference in gains in this grade is .09. The gain of .17 made by the control group is not three times this SD and is therefore from this study not acceptable as evidence of the superiority of the control method in this grade. There is the possibility that the pupils' former problem solving habits may account for this result—as the experience of the pupils increased in the use of a method previously learned there was more and more present the necessity for the first removal of those habits already learned and the accompanying difficulties of adjustment to a new method. The experimental group of pupils had to make a change; the control pupils continued with a method with which they were familiar.

In any consideration of the comparative results in the grades above 3A it may be well to recognize the fact that those results may have beein influenced greatly by that very necessity for a

---

[9] McCall, W. A. *Op. cit.*, page 185.

readjustment in method on the part of the pupils in the experimental group and for the formation of different habits in problem solving. In these grades all the pupils had received problem instruction by the control method for periods of from six to eighteen months, and it may well be supposed that they had acquired certain definite habits in problem attack. The pupils of the experimental group were then required to substitute a new method that may have necessitated a breaking down of habits already learned. During that period of readjustment, indefinite in length, there might be a loss in problem solving ability that would be evidenced in their results until such time as the new method had been so mastered as to guarantee complete functioning. Continued drill and practice in the use of the new method, beyond that possible in the limited time of the experimental period, might have produced that mastery and also results similar to those in the 3A grade where the readjustment was not necessary. It is to be noted that the analysis method, in spite of the difficulties it may have met, produced a steadily increasing gain[10] in number of problems solved through grades 3A to 4A, although the rate of gain was not relatively as great as that produced by the control method. The new analysis method made an *amount* of gain in the 4A grade equal to the established control method that had been in use for a full year. It almost equalled the amount of the control method's gain in the 4B grade where the pupils continued in the use of a method to which they had been accustomed for a period of eighteen months. This difference in rate of gain was probably caused by the increasing difficulty of readjustment on the part of pupils increasingly familiar with a former method. It will remembered that the control method was in use for some time and that the control group pupils had been drilled in it in previous grades as prescribed by the Philadelphia Course of Study in Arithmetic. Those pupils continued with a method that was

---

[10]

Gains by Grades	3A	3B	4A	4B
Experimental	2.04	2.17	2.30	2.02
Control	1.72	2.10	2.30	2.19

possibly well established. They had no new habits to master and were certainly not confused by any necessity for a change in method.

In grades 3B-4B where there may have existed a necessity for removal of habits previously learned by all pupils, the length of the experimental period may have been too short to assure the complete mastery of the analysis method. It would be interesting to know the relative results in all of the grades had the two methods been continued for an entire term. What would be the results should the two methods be continued in these same classes during the next term?

A comparison of these results with the results of the experiment in Reading, Pa. (See page 33) will show a very similar possibility of definite findings in favor of the analysis method in the third grades. Two of the third grades in that city made gains favorable to the analysis method. In the fourth grades, however, the greater gains are again on the side of the control method. It was also doubtless equally true in Reading that there was the same necessity for a change in method on the part of some of the pupils who have had considerable experience in problem solving by a method previously learned and established. That, and the limited time devoted to the experiment, may have produced the similar results.

The first Philadelphia experiment (page 20), although not directly comparable in its results because of differences in experimental procedure, also shows a definite superiority for the analysis method in the third grades. In the 3A grade the difference was quite marked. In this grade the highest grade score was made by the experimental group even though its median score in the first test was considerably lower than the median of the control group. In the other grades, also, the experimental classes made greater *gain* in median scores.

Summary of Conclusions

1. All normal pupils definitely engaged in problem solving profit from specific instruction and drill.

2. In the third grades, where definite instruction in problem

solving usually begins, the analysis method of instruction produced better results during the experimental period.

3. Probably due to the limited time devoted to the experiment which may not have allowed for the complete mastery of the analysis method with the attendant necessity for the breaking down of problem solving habits previously learned, the evidence produced by this study for the fourth grades does not indicate a decided advantage for either of the methods employed.

APPENDIX

PROBLEM TEST—1

1. A fruit man had 36 oranges in a box. He sold 24 of them to Frank's mother. How many oranges were left in the box?

2. How much money must Alice take to the store to buy a 7 cent box of crackers and a can of corn that costs 18 cents?

3. Anna saw a doll in a store window. It cost 80 cents. How much more money does she need to buy it if she has only 62 cents now?

4. How much money will Harry need to buy 4 wheels for his roller skates if the wheels cost 22 cents each?

5. Four boys earned 48 cents selling old newspapers. If they share the money equally, how much will each boy get?

6. How many piles of oranges could a storekeeper make from 54 oranges if he put 6 oranges in each pile?

7. Bob earns 25 cents an hour working at a grocery store. If he should work 7 hours this week, how much money would he receive?

8. Jane's mother bought a 6-pound roast of beef at $.32 a pound. She gave the man $5.00. How much change should she receive?

9. How much change would you receive from a five dollar bill if you bought two baseballs at $1.25 each and a catcher's glove for $2.15?

10. On the first three days of a trip Tom traveled 92 miles, 87 miles, and 97 miles. What was the average distance he traveled each day?

11. Mr. James bought a new car that cost $1050. He made a first payment of $450, and agreed to pay $75 a month until the rest was paid. How long did it take him to pay the entire amount?

12. A farmer raised 120 bushels of potatoes. He sold all but ¼ of them at $1.25 a bushel. How much money did he get?

PROBLEM TEST—2

1. Fred had 28 cents in his bank. He bought a top that cost 10 cents. How much did he have left?

2. Harry wants to buy a copy book that costs 8 cents and a ball for 18 cents. How much money must he have?

3. Frank wants a baseball that costs 75 cents, but he has only 53 cents. How much more money does he need?

4. How much money will Jane have to pay for 3 pounds of nuts that cost 28 cents a pound?

5. Three boys bought a baseball that cost 75 cents. How much should each boy pay if they agreed to share the cost equally?

6. How many days will it take a class to read a story that is 36 pages long, if the class reads 4 pages each day?

7. How much will Fred's mother have to pay for a turkey that weighs 8 pounds and costs 53 cents a pound?

8. How much change should you receive from $2.00 if you buy a 5-pound roast of beef at $.37 a pound?

9. Jack bought a radio battery for $1.75 and 3 radio tubes at $1.35 each. He gave the dealer a ten dollar bill. How much change should he receive?

10. In the tests last week Harry's marks were 85, 92 and 90. What was his average mark for the week?

11. A farmer bought a horse for $165, paying $40 down and the balance in 5 months. How much did he pay each month?

12. A dealer bought 75 dozen oranges that he offered for sale at $.65 a dozen. At the end of the day he had sold all but ⅓ of them. How much did he receive for the oranges sold?

TEST DIRECTIONS

1. This test is to be given to all pupils in grades 3A–4B inclusive.

2. Material—See that each pupil has two sharpened pencils and two sheets of paper of the kind regularly used in the class for arithmetic work.

3. Procedure—Distribute papers face down and tell pupils not to look at them until the signal is given. Say to the pupils, "This is a test in problem solving. Find the answers to as many of these problems as you can, doing all the necessary work on the arithmetic paper I have given you. You are not expected to be able to do all the problems—when you have finished all you can do, then give your papers to me. Start and stop promptly at the signal. Ready—Begin."

4. Time—3A—30 minutes
3B—30 minutes
4A—35 minutes
4B—35 minutes
} Collect *all* papers at the end of the time limit.

5. Marking—Test papers are to be marked only on the basis of correct reasoning. A problem may show the wrong answer but the correct arithmetical processes may have been used—such a problem is to be marked correct. Errors in computation are not to be considered. Place on the pupil's paper the number of problems correctly reasoned.

Illustration—"Find the cost of 8 dozen eggs at $.46 a dozen."

$.46
8
——
$3.58 mark this problem correct because the pupil used multiplication.

Illustration—"If 24 children are placed in 6 rows, how many children will be in each row?"

6
4
—
24

Mark this problem *incorrect.* The pupil has the correct answer but it is a process of reasoning possible only with small numbers.

6. Recording Results—On the Class Record Sheet, opposite the proper pupil score, enter the number of pupils making each score. Median is determined in the usual way.

7. Return the Class Record Sheet to your Principal within one week.

8. Save this sheet of test directions.

EXPERIMENT IN PROBLEM SOLVING

The "Analysis" Method

1. Preliminary lesson:
   (a) Name the four fundamental processes used in arithmetic.
   (b) Use two numbers in each operation:

6	6	6		3
				2) 6
2	2	2		
—	—	—		
8	4	12		

   (c) Two operations give a larger result—addition and multiplication.
   (d) Two operations give a similar result—subtraction and division.
   (e) Drill on above facts for mastery.

The one-step problem: Teaching procedure.

1. Read the entire problem carefully.
2. Determine what is to be found.
3. Select the facts that will help to find the answer.
4. Determine the name of the answer.
5. "Will the answer be more or less than ....?" The teacher will fill in the blank in her question according to the facts of the problem. (See illustration below.)
6. "Which operations will give us (less) (more) for an answer?"
7. "Which shall we use here?"

As an additional aid in making the correct choice of operation the teacher will indicate wherever possible that one of the operations could not be used, viz.: we cannot add apples and cents, etc. Only like things can be added or subtracted.

Illustration—"How many apples will Tom need to fill 4 baskets if he puts 6 apples into each basket?"

1. The problem is read.
2. "What are we asked to find?"
3. "What do we know that will help us find the answer?"—that there are 4 baskets and Tom puts 6 apples into each.
4. "What will be the name of the answer?"—apples.
5. "Will he need more or less than 6 apples?" Select the number in the problem that corresponds to the name of the answer. This device cannot be used in some division problems.
6. "What two operations give us more for an answer?"—addition and multiplication.
7. "Which shall we use here?"—multiplication.

"Why could we not use addition?"—because you cannot add "apples" and "baskets".

The two-step problem: Teaching procedure.

1. 2. 3. Same as in one-step problem.
4. Find out whether every fact necessary to solve the problem is known.
5. Determine the unknown facts.
6. "What will be the name of the first answer that we must get?"
7. "Will it be more or less than .....?"
8. "Which two operations give us (more) (less) for an answer?"
9. "Which shall we use here?"
10. "Will the final answer be more or less than .....?" Fill in the proper facts where possible.
11. "Which operation shall we use here?"
12. "What numbers shall we use?" This question to be asked only when the actual solution is made, by teacher or pupil.

Illustration—"James worked for two hours on one afternoon and for 6 hours on Saturday. If his pay was $.25 an hour, how much did he earn?"

1. The problem is read.
2. "What are we asked to find?"
3. "What do we know that will help us find it?"
4. "Do we know all that we need to know to find the answer?"
5. "What must be known first?"—how many hours he worked in all.
6. "What will be the name of the first answer?"—hours.
7. "Will it be more or less than 5 hours?" .... more
8. "What operations give more for an answer?"
9. "Which shall we do here?"
10. "What will be the name of the final answer?"—money.
11. "Will it be more or less than 25 cents?"—more.
12. "What operations give more for an answer?"—addition and multiplication.

13. "Which shall we use here?"—multiplication.
14. "What numbers shall we use?"—8 and 25.

Procedure

First Problem Lesson of Each Week—Problems placed on blackboard in advance or selected from texts. Consideration of each problem following specified procedure outlined above, work directed by the teacher. No written work or complete solution of any problem in this lesson.

Second Problem Lesson—First problem considered as on preceding day. The second problem is to be completely solved, by the teacher on the board at the direction of the pupils, with emphasis on written form and correct labeling of final answer. The remainder of the period to be spent in written solutions by pupils, teacher giving individual help where necessary.

Last 2 Problem Lessons—Spent entirely in written solutions by the pupils in the manner of test lessons, with no teacher help.

Important

Select problems that require fundamental operations of but normal difficulty. Be sure problems are within the grade limits, of course.

Please do not make any homework assignments in problem solving during the experimental period. Adhere strictly to the time limits set, both in number of lessons and in length. The experiment is a test of methods, *not* of the teachers.

Two-step problems begin in the 3B grade on the ninth lesson (making change only). Fourth grade teachers will begin two-step problems with the fifth lesson.

The "Control" Method

Teachers in charge of Control classes will follow exactly the method outlined in the Philadelphia Course of Study (pp. 82, 103, 131, and 150) and will be very careful not to depart from it in any way nor to extend the suggestions there given by the use of any personal device or method of teaching to which they may have been accustomed. They will be careful to adhere strictly to the number of lessons and the length of lessons as prescribed for *all* classes.

Following the suggestions on page 82 of the Course, the lesson should proceed as follows:

Problem—"How much will Frank have to pay for 3 cans of peas that are sold at the store for 18 cents a can?"

1. Read the problem carefully.
2. Teacher asks, "What are we asked to find?"
3. "What do we know that will help us find it?"

4. "Shall we add, subtract, multiply, or divide?"
5. The solution is then performed.
   a. By the teacher at direction of class—first problem day of each week.
   b. By pupils, at seats and blackboard—second problem day of each week.

The entire solution may then be presented orally by one or more pupils. Any special aids in making the choice of operation, as indicated on page 83, should be taught incidentally and should not take up a large part of the problem lesson.

The two-step problem should be taught by following exactly the method outlined on page 132 of the Course of Study for Grades 4A and 4B.

PROCEDURE

First Problem Day of Each Week—Problems placed on the board in advance or selected from textbook. Consideration of each problem, following the specified procedure, directed by the teacher. Solution of the problems by the teacher at the blackboard, directed by the pupils who state the numbers to be used and the process.

Second Problem Day of Each Week—Consideration of the first problem by the class as above, with solution at blackboard by teacher at direction of pupils. The remaining problems will then be solved by the pupils at the blackboard and at their seats. The teacher will give individual help where necessary, but will always adhere to the regular class method.

Last Two Problem Lessons—To be spent entirely in written solutions by the pupils at their seats, in the manner of test lessons, without help from the teacher.

It is earnestly requested that no homework assignments in problem solving be made during this experiment.

TABLE 25
MEDIAN SCORES OF CLASSES IN NUMBER OF PROBLEMS CORRECTLY REASONED

GRADE 3A CLASS	EXPERIMENTAL TEST 1	EXPERIMENTAL TEST 2	CLASS	CONTROL TEST 1	CONTROL TEST 2
1	4.0	4.4	1a	4.0	5.5
2	3.5	7.1	2a	3.7	6.3
3	1.9	4.1	3a	1.9	3.3
4	2.8	6.7	4a	2.8	3.8
5	2.8	3.9	5a	2.7	5.5
6	2.6	4.2	6a	2.7	4.3
7	2.5	4.6	7a	2.4	3.4
8	2.5	4.2	8a	2.6	4.8
9	2.1	4.2	9a	2.2	3.9
10	3.0	4.2	10a	3.0	3.5
11	1.8	2.5	11a	1.7	3.9
12	2.5	5.4	12a	2.7	5.0
13	2.9	5.4	13a	2.9	4.8
14	4.2	6.7	14a	4.4	5.7
TOTAL	39.1	67.6	TOTAL	39.7	63.7
AVERAGE	2.79	4.83	AVERAGE	2.83	4.55
	Gain 2.04—73.11%			Gain 1.72—60.77%	
Sigma	1.30	1.92	Sigma	1.41	1.73

Table reads: Class 2 correctly reasoned (median score) 3.5 problems in the first test and 7.1 problems in the second test, etc.

GRADE 3B CLASS	EXPERIMENTAL TEST 1	EXPERIMENTAL TEST 2	CLASS	CONTROL TEST 1	CONTROL TEST 2
1	6.4	7.6	1a	6.3	8.2
2	3.7	6.0	2a	3.8	6.9
3	3.6	4.9	3a	3.5	4.4
4	4.4	6.3	4a	4.5	6.0
5	5.0	7.3	5a	4.8	6.8
6	2.2	5.3	6a	2.4	4.6
7	3.9	6.6	7a	3.9	6.9
8	4.2	5.9	8a	4.3	6.5
9	2.7	5.7	9a	2.5	4.6
TOTAL	36.1	55.6	TOTAL	36.0	54.9
Average	4.01	6.18	Average	4.00	6.10
Sigma	1.71	2.00	Sigma	1.72	2.01
	Gain, 2.17—54.11%			Gain, 2.10—52.5%	

Grade 4A	Experimental			Control	
class	test 1	test 2	class	test 1	test 2
1	7.0	10.6	1a	7.0	9.6
2	6.6	8.5	2a	6.6	10.8
3	6.3	8.8	3a	6.2	8.0
4	4.9	8.8	4a	4.9	7.4
5	5.3	6.1	5a	5.3	7.0
6	6.5	8.8	6a	6.9	9.0
7	5.3	8.5	7a	5.3	6.8
8	7.6	10.7	8a	7.4	10.2
9	3.8	5.3	9a	4.0	7.0
10	4.6	6.8	10a	4.8	6.1
11	5.4	6.6	11a	5.6	7.2
12	5.4	7.3	12a	5.2	6.5
13	4.5	6.7	13a	5.1	9.0
14	7.2	9.1	14a	6.0	8.0
Total	80.4	112.6	Total	80.3	112.6
Average	5.74	8.04	Average	5.74	8.04
Gain, 2.30—40.0%			Gain, 2.30—40.0%		
Sigma	2.06	2.63	Sigma	2.00	2.49

Grade 4B	Experimental			Control	
class	test 1	test 2	class	test 1	test 2
1	6.6	6.7	1a	6.6	9.1
2	6.0	9.0	2a	6.1	8.2
3	8.0	8.3	3a	8.0	9.8
4	8.9	11.7	4a	8.9	11.5
5	8.7	10.9	5a	8.5	11.2
6	8.3	10.3	6a	8.2	10.1
7	6.4	8.2	7a	6.8	8.5
8	7.9	11.2	8a	7.7	9.7
9	7.3	7.6	9a	7.4	10.1
10	7.0	8.8	10a	7.6	10.3
11	5.7	10.3	11a	4.6	6.0
Total	80.8	103.0	Total	80.4	104.5
Average	7.34	9.36	Average	7.31	9.50
Sigma	2.32	2.55	Sigma	2.37	2.31
Gain, 2.02—27.52%			Gain, 2.19—29.95%		

EVALUATION OF TESTS A AND B, IN NUMBER OF PROBLEMS CORRECTLY REASONED BY THE 2321 PUPILS IN THE FIRST PHILADELPHIA EXPERIMENT

PROBLEM	TEST A	TEST B	GAIN PER CENT
1	1852	2060	8.9
2	2037	2147	4.5
3	2040	2123	3.5
4	1846	2057	9.0
5	1424	1741	13.6
6	1542	1764	9.5
7	1418	1788	15.9
8	797	1223	18.3
9	975	1367	16.8
10	982	1464	20.7
11	503	1108	26.0
12	496	815	13.7
13	298	569	11.6
14	188	396	8.9
15	249	482	10.0
16	54	266	9.1
Average			12.5

BIBLIOGRAPHY

Boyer, Philip A., "Diagnosis in Problem Solving in Arithmetic," *Sixteenth Annual Schoolmen's Week Proceedings, U. of P.*, 1929, pp. 357–370.

Buswell and Judd, "A Summary of Educational Investigations Relating to Arithmetic," University of Chicago, 1925.

Chase, V. E., "Diagnosis and Treatment of Some Common Difficulties in Solving Verbal Problems in Arithmetic," Bulletin No. 12, Bureau of Research, Fordson, Mich., 1928.

Greene, Harry E., "Directed Drill in the Comprehension of Verbal Problems," *Journal of Educational Research, XI*, January, 1925, pp. 33–40.

Hydle, L. P. and Clapp, F. L., "Elements of Difficulty in the Interpretation of Concrete Problems in Arithmetic," Bureau of Research, University of Wisconsin, No. 9, September, 1927.

Kelley, T. L., "Interpretation of Educational Measurements."

Kelley, T. L., "Statistical Method."

Lessenger, W. E., "Motivation and the Accomplishment Quotient Technique," College of Education Library, University of Iowa, 1925.

Lutes, Olin S., "An Evaluation of Three Techniques for Improving Ability to Solve Arithmetic Problems," University of Iowa, Monograph No. 6, June, 1926.

McCall, W. A., "How to Experiment in Education."

Newcomb, R. S., "Teaching Pupils How to Solve Problems in Arithmetic," *Elementary School Journal*, November, 1922, pp. 183–189.

Osburn, W. J., "Diagnostic and Remedial Treatment for Errors in Arithmetical Reasoning," State Department of Public Instruction, Madison, Wis., mimeographed material.

Stevenson, P. R., "Difficulties in Problem Solving," *Journal of Educational Research, XI*, February, 1925, pp. 95–103.

Terry, Paul W., "The Reading Problem in Arithmetic," *Journal of Educational Psychology, XII*, October, 1921, pp. 365–377.

Washburne and Osborne, "Solving Arithmetic Problems," *Elementary School Journal*, December, 1926, pp. 296–304.

Wilson, Estaline, "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal, XXII*, January, 1922, pp. 380–386.